

# Introduction to R

Chanmoo Park

Department of Statistics  
Seoul National University

EPIC Seminar

February 8, 2017

# What is “R”?

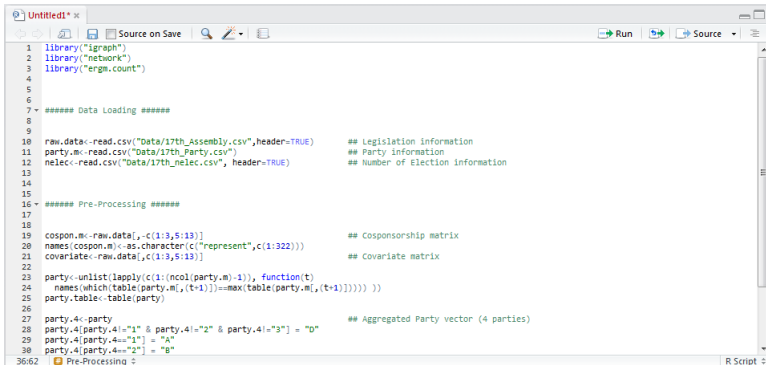
## WHAT IS “R”?



- ① 프로그래밍 “언어”
  - Click, 혹은 Drag and Drop 방식이 아닌, 직접 약속된 명령어를 입력하는 방식
- ② “통계 분석을 목적”으로 하는 언어
- ③ 무료, 오픈 소스 (전 세계 개발자들이 직접 업데이트 가능)

# What is "R"?

## PROGRAMMING LANGUAGE "R"



```
1 library("igraph")
2 library("network")
3 library("ergm.count")
4
5
6
7 ##### Data Loading #####
8
9
10 raw.data<-read.csv("Data/17th_Assembly.csv",header=TRUE)      ## Legislation information
11 party.m<-read.csv("Data/17th_Party.csv")                      ## Party information
12 nelec<-read.csv("Data/17th_nelec.csv", header=TRUE)           ## Number of Election information
13
14
15
16 ##### Pre-Processing #####
17
18
19 cospon.m<-raw.data[,c(1:3,5:13)]                                ## Cosponsorship matrix
20 names(cospon.m)<-as.character(c("represent",c(1:322)))
21 covariate<-raw.data[,c(1:3,5:13)]                              ## Covariate matrix
22
23 party<-unlist(lapply(c(1:(ncol(party.m)-1)), function(t)
24   names(which(table(party.m[, (t+1)]==max(table(party.m[, (t+1)]))))))
25 party.table<-table(party)
26
27 party.4<-party
28 party.4[party.4!="1" & party.4!="2" & party.4!="3"] = "p"
29 party.4[party.4=="1"] = "A"
30 party.4[party.4=="2"] = "B"
```

36/62 Pre-Processing R Script

# What is “R”?

## PROGRAMMING LANGUAGE “R”

```

1 library("igraph")
2 library("network")
3 library("ergm.count")
4
5
6
7 ##### Data Loading #####
8
9
10
11 party.m<-read.csv("Data/1759_Party.csv",header=T)
12
13
14
15
16 ##### Data Processing #####
17
18
19 cospon.m<-raw.data[,c(1:2,5:6)]
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62

```

- 비교적 “자유로운” 분석이 가능하다
- 분석 결과의 출력물을 다양하게 편집할 수 있다.
- 연구의 재현(Reproducing)이 쉽다.
- 언어에 대한 문법과, 알고리즘에 대한 간단한 이해가 필요하다.
- “그나마” 다른 언어(C++, python 등)에 비해서 직관적이다.
- 프로그래밍 언어이기 때문에 빠를 것 같지만, 그렇지 않다.

# What is “R”?

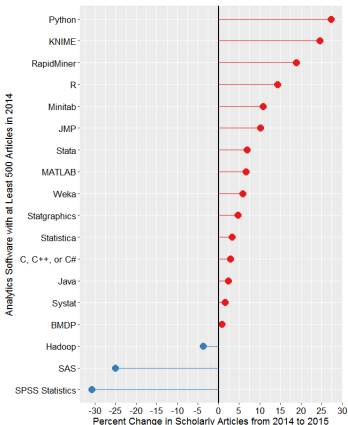
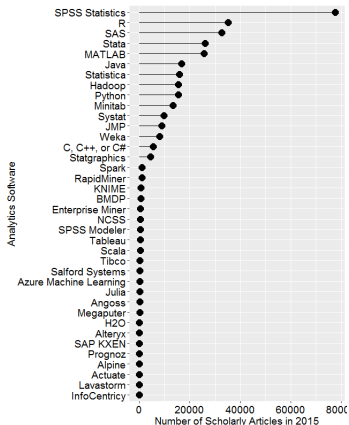
## STATISTICAL LANGUAGE “R”

TYPE OF STATISTICAL ANALYSIS	R	MATLAB	SAS	STATA	SPSS
Nonparametric Tests	Yes	Yes	Yes	Yes	Yes
T-test	Yes	Yes	Yes	Yes	Yes
ANOVA & MANOVA	Yes	Yes	Yes	Yes	Yes
ANCOVA & MANCOVA	Yes	Yes	Yes	Yes	Yes
Linear Regression	Yes	Yes	Yes	Yes	Yes
Generalized Least Squares	Yes	Yes	Yes	Yes	Yes
Ridge Regression	Yes	Yes	Yes	Limited	Limited
⋮	⋮	⋮	⋮	⋮	⋮
Total(57)	57	57	42	29	20
Popular methods(Frequently Used 20)	20	20	19	16	15

**Table:** Table comparing the statistical capabilities of software packages  
[http://stanfordphd.com/Statistical\\_Software.html](http://stanfordphd.com/Statistical_Software.html))

# What is “R”?

## STATISTICAL LANGUAGE “R”



# What is “R”?

## FREE & OPEN SOURCE LANGUAGE “R”

- R로 할 수 있는 모든 작업이 무료이다.
- 사용자들이 직접 유용한 도구들(R 패키지)를 제작하여 공유할 수 있다.
- 하지만 위와 같은 특성상 R의 모든 출력결과에 대해 아무도 공식적으로 보증해주지는 못한다.
- 많은 이용자들이 이용하기 때문에, 정보를 얻기 매우 쉽다.  
(“xxx in R”)

## "R"? or Others?

- 단순 통계 분석, SPSS나 SAS로 구현이 가능한 경우  
→ 익숙한 software
- SPSS나 SAS로 재현 불가능한 통계적 분석  
→ "R" 이용
- 익숙한 software에서 원하는 분석의 세부적인 결과나 효과적인 시각화를 제공하지 않는 경우  
→ "R" 이용



# How to Learn “R”

- Official R Tutorial  
<https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>
- R with basic statistical knowledge  
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 6). New York: springer.
- Beautiful Visualization  
Winston Chang. (2012). *R Graphics Cookbook*. O'Reilly Media.
- “xxx in R” @Google